



Article A Monitoring Method Based on FDALM and Its Application in the Sintering Process of Ternary Cathode Material

Ning Chen 💩, Fuhai Hu, Jiayao Chen *💩, Kai Wang, Chunhua Yang and Weihua Gui

School of Automation, Central South University, Changsha 410083, China

* Correspondence: jychen@csu.edu.cn

Abstract: In industrial processes, the composition of raw material and the production environment are complex and changeable, which makes the production process have multiple steady states. In this situation, it is difficult for the traditional single-mode monitoring methods to accurately detect the process abnormalities. To this end, a multimode monitoring method based on the factor dynamic autoregressive hidden variable model (FDALM) for industrial processes is proposed in this paper. First, an improved affine propagation clustering algorithm to learn the model modal factors is adopted, and the FDALM is constructed by combining multiple high-order hidden state Markov chains through the factor modeling technology. Secondly, a fusion algorithm based on Bayesian filtering, smoothing, and expectation-maximization is adopted to identify model parameters. The Lagrange multiplier formula is additionally constructed to update the factor coefficients by using the factor constraints in the solving. Moreover, the online Bayesian inference is adopted to fuse the information of different factor modes and obtain the fault posterior probability, which can improve the overall monitoring effect of the model. Finally, the proposed method is applied in the sintering process of ternary cathode material. The results show that the fault detection rate and false alarm rate of this method are improved obviously compared with the traditional methods.

Keywords: multimodality; factor modeling; process monitoring; FDALM

1. Introduction

Process monitoring plays an important role in ensuring product safety, improving product quality, and reducing production costs. Therefore, it is an important way to ensure the safe and efficient operation of the production process [1–3]. Multivariate statistical process monitoring (MSPM) has become increasingly popular with the proliferation of digital factories and the massive collection of production process data. The traditional representative techniques include Principal Component Analysis (PCA) and Partial Least Squares (PLS), which are used to monitor chemical production processes and achieve results [4–6].

However, the dynamics information contained in sequence data has been greatly ignored by most process monitoring [7]. Scholars have proposed dynamic augmented matrix [8], autoregressive [9], and linear dynamic systems [10] to monitor the dynamic process. However, the augmented matrix increases the curse of dimensionality. Moreover, the order of the autoregressive model and the dimension of the latent state is difficult to determine, resulting in limited monitoring performance [11]. Fortunately, linear dynamic systems, as a dynamic latent variable model, are good at describing the interaction between first-order dynamic process states. Aiming at the high-order dynamic process [12], Chen et al. [13] proposed a monitoring method of the dynamic autoregressive latent variable model to monitor the preparation process of ternary cathode material. However, the model assumes that the system operates stably under a single operating condition and fails for multimodal processes [14].



Citation: Chen, N.; Hu, F.; Chen, J.; Wang, K.; Yang, C.; Gui, W. A Monitoring Method Based on FDALM and Its Application in the Sintering Process of Ternary Cathode Material. *Sensors* **2022**, 22, 7203. https://doi.org/10.3390/s22197203

Academic Editors: Yuan Yao and Ruben Puche-Panadero

Received: 24 July 2022 Accepted: 15 September 2022 Published: 22 September 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/).

Researchers have made several contributions to addressing the multimodal problem [15,16]. A traditional approach is to build different monitoring models for various operating modes. Zhao et al. [17] proposed a mixed principal component analysis model for monitoring multimodal processes. It established PCA models for multiple data sets to distinguish different modal information. At the same time, Zhao et al. [18,19] proposed methods of PCA and multiple PLS models to handle multimodal processes. This type of method has several drawbacks: not fusing the correlation information between different models; splitting the useful information hidden between the data sequences; and being highly dependent on the similarity measurement algorithm. To address such problems, Bayesian frameworks and combinatorial strategies have been adopted. Yu et al. [20,21] proposed a monitoring method for the Gaussian mixture model and used the Bayesian inference method to fuse various model indicators. Ge et al. [22] proposed a hybrid PPCA model combined with Bayesian inference technology for process monitoring. The above methods attempt to construct a unified mathematical model, and then use Bayesian inference techniques to fuse the difference information between the models. However, they do not consider the information between the data sequences in the transition period. To this end, Wang et al. [23] proposed a Hidden Markov Model-based fault monitoring method for multi-mode processes, which relies on the accuracy of the partitioning algorithm for steady-state and transition-state data sets.

The multi-mode process monitoring method proposed above highly depends on the similarity measurement algorithm to categorize the data samples into a certain group, while ignoring the state action relationship between sequences. Due to the above reasons, the samples during the mode switching cannot be effectively considered, which may increase the error detection rate caused by misclassification [24]. To make full use of the information between sequences and preserve the interaction between multimodal states, a multimodal process monitoring method based on the factor dynamic autoregressive latent variable model is proposed in this paper. First, from a data perspective, a factor dynamic autoregressive latent variable model, which is obtained by combining multiple high-order hidden state Markov chains through factor modeling technology, is constructed. Then, the state output of the sample in each mode is fused to the posterior failure probability of the sample through Bayesian inference online, so as to take into the overall monitoring effect of the multimodal model. The main contributions can be summarized as: (1) with the help of factor modeling technology, multiple high-order Markov state chains are combined to obtain FDALM to simulate the dynamic evolution of multimodal processes; (2) a parameter identification algorithm is proposed to learn the FDALM parameters; (3) Bayesian technology is used to fuse different sub-modal features into sample posterior failure probability to achieve process monitoring tasks and reduce the false alarm rate; (4) focuses on solving the multimodal problems in the sintering process of ternary cathode materials. The proposed method is applied to the sintering process of ternary cathode material manufacturing. The results have practical guiding significance.

The main structure of the paper is arranged as follows. Section 2 introduces the difficulties in monitoring the preparation process of ternary cathode material. Section 3 presents the modeling method of the factor dynamic autoregressive latent variable model and introduces the parameter identification algorithm of the model in detail. A monitoring method based on FDALM is proposed in Section 4. Section 5 applies the monitoring method to monitor the sintering process of ternary cathode material to verify the effectiveness of the proposed method. The last section provides the conclusion.

2. Description of the Problem in Monitoring the Preparation Process of Ternary Cathode Material

The roller kiln, which has a total length of more than 40 m and consists of 21 temperature zones, is a device used for the sintering reaction of ternary cathode material. It is divided into three temperature sections according to the temperature trend: the heating section, the constant temperature section, and the cooling section. Each temperature section is spliced together by several temperature zones, and its cross-sectional view is shown in Figure 1. The material is loaded into the saggar, enters the replacement chamber at the entrance of the roller kiln, and starts a series of chemical reactions. The qualified product relies on the temperature of the internal temperature zone of the roller kiln, the transmission power, and the supply of oxygen. The heating system adjusts the temperature field distribution in the temperature zone by controlling the current flowing on the silicon carbide rod, so as to ensure the uniform heat conduction of the material during the transmission process. The transmission system is driven by the rollers moving at a constant speed to help the saggars placed on the rollers to move forward smoothly and slowly at a certain speed. Oxygen is introduced into the air inlet to provide oxygen for the compound reaction, and the exhaust gas is extracted from the reaction. This system further ensures the stable air pressure condition in the roller kiln. In the ternary cathode material roller kiln, the temperature is mainly measured by thermocouples. Two thermocouples are at the top and bottom of each temperature zone, close to the upper and lower silicon carbon rods.



Figure 1. Sectional view of a roller kiln.

Due to the limitation of the production environment, there is a certain hysteresis relationship between the control parameters and the sintering state, so it is impossible to directly evaluate the quality of the production and preparation process by controlling variables such as silicon carbide rod current and oxygen concentration. At the same time, the quality of the ternary material is related to the temperature of the entire process, and the collected temperature variables have high dimensions, so it is difficult to measure whether the process state is in a normal state through a simple weighted average index. In addition, in order to compensate for the influence of environmental factors, there is more than one optimal sintering system for the sintering preparation process [25]. For example, the operating parameters in winter and summer are different, so a single evaluation criterion cannot be applied to the preparation process of ternary material. These problems make it difficult for operators to accurately grasp the state of on-site sintering preparation in real-time, resulting in side effects such as lag in operation and fluctuations in product quality. Therefore, it is very important to study the process monitoring technology to fully analyze the data collected by the process sensor to guide the operator to grasp and adjust the operating parameters in time, so that the process can return to its normal state.

That is to say, a qualified battery product requires that the distribution of the internal temperature field of the roller kiln conforms to a certain curve, that is, the temperature trend is shown in Figure 2. In Figure 2, 1 is the furnace exhaust port, 2 is the upper silicon-carbon rod, 3 is the furnace air inlet, and 4 is the lower heated silicon-carbon rod. Since the process conditions are directly determined by the sintering temperature, the temperatures in every section and product quality need to be modeled and monitored. The product quality is related to the temperatures of the entire section, and the front and rear temperature zones affect each other, so the monitoring of a single temperature zone cannot achieve the overall monitoring goal. However, thermocouples are installed at the top and bottom of each temperature zone inside the sintering device as temperature measuring elements

to measure the temperature of the kiln. These temperatures can indirectly determine the quality of the products in the current kiln as a whole. At the same time, to monitor the faults related to product quality, such as the abnormal increase of residual lithium in the outlet replacement chamber, it is necessary to increase the monitoring of product quality based on monitoring temperature variables. The indicator that can best reflect product performance is surface free radicals. When the free lithium content increases, the material easily absorbs carbon dioxide and moisture in the air to generate lithium carbonate and lithium hydroxide, which not only reduces the storage stability of the material in the air but also hinders the deintercalation and electron conduction of lithium ions in the raw material.



Figure 2. Variation trend of the internal temperature of a roller kiln.

The curve in Figure 2 records the temperature distribution inside the kiln that needs to be achieved under the production of qualified products. The actual kiln temperature is affected by the external environment, and the controller will eliminate these effects to adjust the temperature profile inside the kiln. Therefore, the sintering regime that deviates from the curve in Figure 2 is also consistent with the normal production process and cannot be judged as a faulty working condition. This paper investigates the use of mathematical models to track these normal operating conditions, and the models themselves are trained using normal data to obtain them. For abnormal data, the state variables of the model deviate from the normal distribution and the process is determined to be abnormal. However, the method in this paper will only monitor the abnormalities but not eliminate them, and the job of eliminating them is left to the fault diagnosis.

After analyzing the data characteristics of the ternary cathode material preparation industrial system, it is found that the production data has problems such as noise interference, dynamics, high dimension, and multi-modality [26]. First, in the sintering preparation process, the uncertainty of the production process such as the aging of the silicon carbide rod and the fluctuation of the exhaust pipe flow causes the acquisition device to be randomly interfered with by factors such as random errors and human errors. Secondly, due to the mutual coupling between the reactions of the device before and after, the samples at the current sampling time are affected by the state of the previous time and show strong dynamic characteristics. Third, because the sintering process has the characteristics of a long process and large delay, the sensors installed in the system will collect numerous variables to obtain the sintering state, resulting in a high dimension of data that is involved. Fourth, the chemical reaction that occurs during the sintering process is particularly sensitive to the temperature of the external environment [27]. When there are periodic disturbances in the external environment, such as seasonal changes, the system will adjust the operating parameters to compensate for the influence of external disturbances. The system has a variety of stable working states, that is, the distribution of the collected data will change periodically with the change in the external environment. To sum up, the sintering process data of ternary cathode material exhibit randomness, dynamics, high latitude, and multimodal characteristics. Traditional monitoring and monitoring methods cannot overcome these characteristics to satisfy the monitoring effect required by the industry.

3. Factor Dynamic Autoregressive Latent Variable Model

In view of the multimode characteristics of complex production process data, this section proposes a factor dynamic autoregressive latent variable model (FDALM) from the data perspective. The technique generalizes a single higher-order linear dynamic system to a multimodal-adaptive process for modeling multimodal features. Subsequently, an improved EM algorithm is proposed for model parameter identification to ensure the inter-class distances of different models.

3.1. Model Structure

Most of the traditional data modeling methods are based on the assumption of a single-mode and a static process, which makes the monitoring method not capable of multimode tracking, resulting in false alarms, misdiagnoses, and even failures in monitoring results. Therefore. for multimode process data modeling, the modal representation is very important. The division between different modes requires a comprehensive trade-off between the dispersion and ductility of each mode, that is, the modes cannot overlap excessively, and the description of the modes is effective and sufficient. In order to make full use of the effective transition information between data series, the adopted factor modeling method is used to extend the single-modal DALM to obtain the modeling method of factor dynamic autoregressive latent variable model (FDALM). The probability topology of the factor dynamic autoregressive latent variable model is shown in Figure 3.



Figure 3. Topology of FDALM.

Multi-modal processes are also generally dynamic. On the one hand, when the system is stable in a certain mode, the states at different times will interact with each other. In addition, the switching between different modes is also a dynamic process. Therefore, whether it is data modeling or modal partitioning, dynamic information cannot be ignored. It can be seen from the above topology diagram that when there are *K* modes in the system, each mode is simulated by a high-order Markov state chain, which fully simulates the dynamic evolution law of a single mode. At the same time, *K* modalities exist at each moment, so the output observation variable is weighted with a weight coefficient, and each state is weighted with a factor weight P(k) to form the observation variable **x** at the current time. The mathematical structure of FDALM is shown in Equation (1).

$$\mathbf{z}_{t,k} = \mathbf{A}_k \mathbf{h}_{t-1,k} + \mathbf{\eta}_k^{\mathbf{z}}$$

$$\mathbf{x}_{t,k} = \mathbf{B}_k \mathbf{z}_{t,k} + \mathbf{\eta}_k^{\mathbf{x}}$$

$$\mathbf{x}_t = \sum_{k=1}^{K} P(k) \mathbf{x}_{t,k}$$
(1)

where $\mathbf{z}_{t,k} \in \mathbb{R}^d$ represents the latent variable describing the system at time *t* in mode *k*, $\mathbf{h}_{t,k}$ represents the augmented latent variable composed of latent variables in mode *k* at time

t and the first L - 1 moments, namely $\mathbf{h}_{t-1,k} = [\mathbf{z}_{t-1,k}^T \ \mathbf{z}_{t-2,k}^T \ \cdots \ \mathbf{z}_{t-L,k}^T]^T \in \mathbb{R}^{dL}$, where *L* is the dynamic order. Then identify the order coefficients learned by the algorithm. $\mathbf{x}_{t,k} \in \mathbb{R}^v$ represents the process observation variable composed of the process variable and the quality variable in the mode *k* at time *t*, and *d* and *v* represent the dimensions of the latent variable and the observed variable, respectively. $\mathbf{A}_k \in \mathbb{R}^{d \times dL}$ represents the state transition matrix from the augmented latent variable $\mathbf{h}_{t-1,k}$ to the latent variable $\mathbf{z}_{t,k}$ at the current moment. $\mathbf{B}_k \in \mathbb{R}^{v \times d}$ represents the divergence matrix of the latent variable $\mathbf{z}_{t,k}$ to the observed variable $\mathbf{x}_{t,k}$. $\mathbf{\eta}_k^z \in \mathbb{R}^d$ and $\mathbf{\eta}_k^x \in \mathbb{R}^v$ are the Gaussian noises of the latent variable and observed variable in mode *k*, respectively. P(k) denotes the prior probability of the *k*th modality, $P(k|\mathbf{x}_i)$ denotes the posterior probability of the *k*th modality under \mathbf{x}_i conditions; \mathbf{x}_t denotes the data collected at moment t, i.e., for the roller kiln temperature variable of this object. Assuming that the noises are independent of each other, the distributions obeyed are $\mathbf{\eta}_k^z \sim N(0, \mathbf{\Sigma}_{\mathbf{x},k})$ and $\mathbf{\eta}_k^x \sim N(0, \mathbf{\Sigma}_{\mathbf{x},k})$, respectively. The parameter $\mathbf{\Theta}$ of the whole factor dynamic autoregressive latent variable model is shown in Equation (2).

$$\boldsymbol{\Theta} = \left\{ \mathbf{A}_{k}, \mathbf{B}_{k}, \boldsymbol{\mu}_{\pi,k}, \boldsymbol{\Sigma}_{\pi,k}, \boldsymbol{\Sigma}_{\mathbf{z},k}, \boldsymbol{\Sigma}_{\mathbf{x},k}, P(k) \middle| k = 1, 2, \cdots, K \right\}.$$
(2)

Among them, $\mu_{\pi,k}$ and $\Sigma_{\pi,k}$ are the mean and variance of the Gaussian distribution of the augmented latent variable $\mathbf{h}_{0,k}$ at the initial time, respectively.

3.2. FDALM Parameter Identification

Compared with DALM parameter identification, the factor modeling method not only needs to ensure the maximum likelihood function of a single model, but also needs to ensure that different model structures do not overlap excessively. Due to the existence of two types of unobservable variables in FDALM structure, which are the latent variables of the model and the factor coefficients P(k), the traditional maximum likelihood function cannot be used. In this section, an improved EM algorithm [28] is proposed to identify the parameters, which shortens the learning time and ensures the inter-class distance. Similar to each single DALM model, in the *k*th sub-DALM, the conditional probability distribution of the latent variable $\mathbf{z}_{t,k}$ and the observed variable $\mathbf{x}_{t,k}$ is shown in Equation (3).

$$\begin{aligned} &\mathbf{h}_{0,k} \sim N(\mathbf{\mu}_{\pi,k}, \mathbf{\Sigma}_{\pi,k}) \\ &\mathbf{z}_{t,k} \Big| \mathbf{h}_{t-1,k} \sim N(\mathbf{A}_k \mathbf{h}_{t-1,k}, \mathbf{\Sigma}_{\mathbf{z},k}) \\ &\mathbf{x}_{t,k} \Big| \mathbf{z}_{t,k} \sim N(\mathbf{B}_k \mathbf{z}_{t,k}, \mathbf{\Sigma}_{\mathbf{x},k}) \end{aligned}$$
(3)

The parameters that the model needs to identify are $\Theta = \{\mathbf{A}_k, \mathbf{B}_k, \boldsymbol{\mu}_{\pi,k}, \boldsymbol{\Sigma}_{\pi,k}, \boldsymbol{\Sigma}_{\mathbf{x},k}, P(k) | k = 1, 2, \dots, K\}$, where $\mathbf{A}_k = [\mathbf{A}_{1,k}, \mathbf{A}_{2,k}, \dots, \mathbf{A}_{L,k}]$. The goal of identification is to obtain the optimal parameter set of the model by giving the observation data matrix data set $\mathbf{x}_{1:T} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_T]^T \in \mathbb{R}^{T \times v}$, where *T* is the number of samples in the training data set. The optimal solution of the model parameters is maximized by solving the log-likelihood function of the model, as shown in Equation (4).

$$\Theta^{opt} = \operatorname{argmax}_{\Theta} L(\Theta) = \operatorname{argmax}_{\Theta} In(\mathbf{x}_{1:T}|\Theta)$$
(4)

In the E step of the EM algorithm, the objective cost function is first constructed, that is, the mathematical expectation of the conditional posterior probability $P(\mathbf{z}_{1-L:T}, k | \mathbf{x}_{1:T}, \Theta^{old})$ of the complete log-likelihood function $InP(\mathbf{x}_{1:T}, \mathbf{z}_{1-L:T} | \Theta)$ about the latent variable is calculated, as shown in Equation (5).

$$Q\left(\boldsymbol{\Theta}\middle|\boldsymbol{\Theta}^{old}\right) = E(L(\boldsymbol{\Theta})) = E\left\{InP\left(\mathbf{x}_{1:T}\middle|\boldsymbol{\Theta}^{old}\right)\right\}$$
$$= \sum_{k=1}^{K} E_{P(\mathbf{z}_{1-L:T},k|\mathbf{x}_{1:T},\boldsymbol{\Theta}^{old})}\left[In\left(\mathbf{x}_{1:T},\mathbf{z}_{1-L:T},k\middle|\boldsymbol{\Theta}^{old}\right)\right]$$
$$= \sum_{k=1}^{K} P(k|\mathbf{x}_{1:T})E_{P(\mathbf{z}_{1-L:T}|\mathbf{x}_{1:T},k,\boldsymbol{\Theta}^{old})}\left[In\left(\mathbf{x}_{1:T},\mathbf{z}_{1-L:T},k\middle|\boldsymbol{\Theta}^{old}\right)\right]$$
(5)

From the Bayesian topology knowledge, combined with the DALM topology, Equation (5) is expanded.

$$InP(\mathbf{x}_{1:T}, \mathbf{z}_{1-L:T}, k) = In\{P(k)P(\mathbf{x}_{1:T}, \mathbf{z}_{1-L:T}|k)\}$$

= InP(k) + InP($\mathbf{h}_{0,k}|k$) + $\sum_{t=1}^{T} \{InP(\mathbf{z}_{t,k}|\mathbf{h}_{t-1,k}) + InP(\mathbf{x}_{t}|\mathbf{z}_{t,k}, k)\}$ (6)

Among them, P(k) is the factor coefficient and obeys $\sum_{k=1}^{K} P(k) = 1$. For the sake of writing convenience, this section will denote the conditional expectation $E_{\mathbf{z}_{1-L:T,k}|\mathbf{x}_{1:T},k,\mathbf{\Theta}^{old}} \{f(\mathbf{x}_t, \mathbf{z}_{t,k})\}$ on the latent variable as $E_Z \{f(\mathbf{x}_t, \mathbf{z}_{t,k})\}$. The mathematical analysis formula of the objective cost function expansion is obtained as Equation (7).

$$Q\left(\boldsymbol{\Theta}|\boldsymbol{\Theta}^{old}\right) = \sum_{k=1}^{K} P(k|\mathbf{x}_{1:T}) \times InP(k)$$

$$-\sum_{k=1}^{K} \frac{P(k|\mathbf{x}_{1:T})}{2} \times \left\{ In|\boldsymbol{\Sigma}_{\pi,k}| + E_{\mathbf{Z}}\left(\mathbf{h}_{0,k}\boldsymbol{\Sigma}_{\pi,k}^{-1}\mathbf{h}_{0,k}^{T}\right) - 2E_{\mathbf{Z}}\left(\mathbf{h}_{0,k}\right)\boldsymbol{\Sigma}_{\pi,k}^{-1}\boldsymbol{\mu}_{\pi,k} + \boldsymbol{\mu}_{\pi,k}^{T}\boldsymbol{\Sigma}_{\pi,k}^{-1}\boldsymbol{\mu}_{\pi,k}\right\}$$

$$-\sum_{t=1}^{T} \sum_{k=1}^{K} \frac{P(k|\mathbf{x}_{t})}{2} \times \left\{ \begin{array}{c} In|\boldsymbol{\Sigma}_{\mathbf{z},k}| + E_{\mathbf{Z}}\left(\mathbf{z}_{t,k}^{T}\boldsymbol{\Sigma}_{\mathbf{z},k}^{-1}\mathbf{z}_{t,k}\right) - 2E_{\mathbf{Z}}\left(\mathbf{h}_{t-1,k}\mathbf{A}_{k}^{T}\boldsymbol{\Sigma}_{\mathbf{z},k}^{-1}\mathbf{z}_{t,k}\right) \\ + E_{\mathbf{Z}}\left(\mathbf{h}_{t-1,k}\mathbf{A}_{k}^{T}\boldsymbol{\Sigma}_{\mathbf{z},k}^{-1}\mathbf{A}_{k}\mathbf{h}_{t-1,k}^{T}\right) \\ In|\boldsymbol{\Sigma}_{\mathbf{x},k}| + \mathbf{x}_{t}^{T}\boldsymbol{\Sigma}_{\mathbf{x},k}^{-1}\mathbf{x}_{t} - 2E_{\mathbf{Z}}\left(\mathbf{z}_{t,k}^{T}\right)\mathbf{B}_{k}^{T}\boldsymbol{\Sigma}_{\mathbf{x},k}^{-1}\mathbf{x}_{t,k} \\ + E_{\mathbf{Z}}\left(\mathbf{z}_{t,k}^{T}\mathbf{B}_{k}^{T}\boldsymbol{\Sigma}_{\mathbf{x},k}^{-1}\mathbf{B}_{k}\mathbf{z}_{t,k}\right) \\ + E_{\mathbf{Z}}\left(\mathbf{z}_{t,k}^{T}\mathbf{B}_{k}^{T}\boldsymbol{\Sigma}_{\mathbf{x},k}^{-1}\mathbf{B}_{k}\mathbf{z}_{t,k}\right) \\ \right\} + cons$$

In the E step of the EM algorithm, given the iterative parameter Θ^{old} of the previous step, the mathematical expectation of the conditional probability of two types of latent variables is obtained. One is the conditional expectation of the latent variable, namely $E_{\mathbf{Z}}(\mathbf{z}_{t,k})$, $E_{\mathbf{z}}(\mathbf{z}_{t,k}\mathbf{z}_{t,k}^T)$, and $E_{\mathbf{Z}}(\mathbf{z}_{t,k}\mathbf{z}_{t-i,k}^T)$, where $t = 0, 1, \dots, T, i = 1, 2, \dots, L$. The derivation process is similar to the E-step of the traditional DALM model, and will not be introduced due to limited space. The final result is given as Equation (8).

$$\begin{cases}
E_{\mathbf{Z}}(\mathbf{z}_{t,k}) = E\left(\mathbf{z}_{t,k} \middle| \mathbf{x}_{1:T}, \mathbf{\Theta}^{old}\right) = \mathbf{m}_{t,k}^{1} \\
E_{\mathbf{Z}}\left(\mathbf{z}_{t,k} \mathbf{z}_{t,k}^{T}\right) = E\left(\mathbf{z}_{t,k} \mathbf{z}_{t,k}^{T} \middle| \mathbf{x}_{1:T}, \mathbf{\Theta}^{old}\right) = \mathbf{M}_{t,k}^{11} + \mathbf{m}_{t,k}^{1}\left(\mathbf{m}_{t,k}^{1}\right)^{T} \\
E_{\mathbf{Z}}\left(\mathbf{z}_{t,k} \mathbf{z}_{t-i,k}^{T}\right) = E\left(\mathbf{z}_{t,k} \mathbf{z}_{t-i,k}^{T} \middle| \mathbf{x}_{1:T}, \mathbf{\Theta}^{old}\right) = \mathbf{M}_{t,k}^{1(i+1)} + \mathbf{m}_{t,k}^{1}\left(\mathbf{m}_{t,k}^{(i+1)}\right)^{T} \\
E_{\mathbf{Z}}\left(\mathbf{z}_{t,k} \mathbf{z}_{t-L,k}^{T}\right) = E\left(\mathbf{z}_{t,k} \mathbf{z}_{t-i,k}^{T} \middle| \mathbf{x}_{1:T}, \mathbf{\Theta}^{old}\right) = \sum_{i=1}^{L} \mathbf{A}_{i,k}\left(\mathbf{M}_{t-1,k}^{iL} + \mathbf{m}_{t-1,k}^{i}\left(\mathbf{m}_{t-1,k}^{L}\right)^{T}\right)
\end{cases}$$
(8)

Among them, $\mathbf{m}_{t,k}$ and $\mathbf{M}_{t,k}$ are the mean and variance of the posterior probability distribution of the *k*th sub-modal latent variable $\mathbf{h}_{t,k}$ with respect to the observation sequence $\mathbf{x}_{1:T}$, respectively. Another type of latent variable is the posterior probability $P(k|\mathbf{x}_t)$ of factor *k*, where $t = 0, 1, \dots, T$. The mathematical solution is shown in Equation (9).

$$P(k|\mathbf{x}_t) = \frac{P(\mathbf{x}_t|k)P(k)}{P(\mathbf{x}_t)}.$$
(9)

Among them, P(k) is the factor coefficient of the previous iteration. In the M step of the EM algorithm, the full log-likelihood expectation function $Q(\Theta | \Theta^{old})$ is maximized by using the latent variable expectation and the posterior probability of the factor obtained in the E step, and a new iterative parameter set Θ^{new} is obtained. First, we maximize the objective cost function $Q(\Theta | \Theta^{old})$, as shown in Equation (10).

$$\boldsymbol{\Theta}^{new} = \operatorname*{argmax}_{\boldsymbol{\Theta}} Q(\boldsymbol{\Theta} | \boldsymbol{\Theta}_{old}). \tag{10}$$

Let the cost function $Q(\Theta|\Theta^{old})$ be derived separately from the model parameters and equal to zero, as shown in Equation (11).

$$\frac{\partial Q(\boldsymbol{\Theta}|\boldsymbol{\Theta}_{old})}{\partial \boldsymbol{\Theta}} = 0. \tag{11}$$

The detailed update process is shown as the following.

$$\frac{\partial \{Q(\boldsymbol{\Theta}|\boldsymbol{\Theta}^{old})\}}{\partial \boldsymbol{\mu}_{\pi,k}} = 0 \Rightarrow \frac{\partial \{2E_{\mathbf{Z}}(\mathbf{h}_{0,k})\boldsymbol{\Sigma}_{\pi,k}^{-1}\boldsymbol{\mu}_{\pi,k} + \boldsymbol{\mu}_{\pi,k}^{T}\boldsymbol{\Sigma}_{\pi,k}^{-1}\boldsymbol{\mu}_{\pi,k}\}}{\partial \boldsymbol{\mu}_{\pi,k}} = 0 \qquad (12)$$
$$\Rightarrow -2\boldsymbol{\Sigma}_{\pi,k}^{-1}E_{\mathbf{Z}}(\mathbf{h}_{0,k}) + 2\boldsymbol{\Sigma}_{\pi,k}^{-1}\boldsymbol{\mu}_{\pi,k} = 0 \Rightarrow \boldsymbol{\mu}_{\pi,k}^{new} = E_{\mathbf{Z}}(\mathbf{h}_{0,k})$$

$$\frac{\partial \{ \mathcal{Q}(\boldsymbol{\Theta} | \boldsymbol{\Theta}^{old}) \}}{\partial \boldsymbol{\Sigma}_{\pi,k}} = 0$$

$$\Rightarrow \frac{\partial \{ In | \boldsymbol{\Sigma}_{\pi,k} | + E_{\mathbf{Z}}(\mathbf{h}_{0,k} \boldsymbol{\Sigma}_{\pi,k}^{-1} \mathbf{h}_{0,k}^{T}) - 2E_{\mathbf{Z}}(\mathbf{h}_{0,k}) \boldsymbol{\Sigma}_{\pi,k}^{-1} \boldsymbol{\mu}_{\pi,k} + \boldsymbol{\mu}_{\pi,k}^{T} \boldsymbol{\Sigma}_{\pi,k}^{-1} \boldsymbol{\mu}_{\pi,k} \}}{\partial \boldsymbol{\Sigma}_{\pi,k}} = 0$$

$$\Rightarrow \boldsymbol{\Sigma}_{\pi,k}^{-1} - \boldsymbol{\Sigma}_{\pi,k}^{-1} E_{\mathbf{Z}}(\mathbf{h}_{0,k} \mathbf{h}_{0,k}^{T}) \boldsymbol{\Sigma}_{\pi,k}^{-1} + \boldsymbol{\Sigma}_{\pi,k}^{-1} E_{\mathbf{Z}}(\mathbf{h}_{0,k}) E(\mathbf{h}_{0,k}^{T}) \boldsymbol{\Sigma}_{\pi,k}^{-1} = 0$$

$$\Rightarrow \boldsymbol{\Sigma}_{\pi,k}^{new} = E_{\mathbf{Z}}(\mathbf{h}_{0,k} \mathbf{h}_{0,k}^{T}) - E_{\mathbf{Z}}(\mathbf{h}_{0,k}) E(\mathbf{h}_{0,k}^{T})$$
(13)

$$\frac{\partial \{ Q(\boldsymbol{\Theta} | \boldsymbol{\Theta}^{old}) \}}{\partial \mathbf{A}_{k}} = 0$$

$$\Rightarrow -\sum_{t=1}^{T} P(k|\mathbf{x}_{t}) \boldsymbol{\Sigma}_{\mathbf{z},k}^{-1} E_{\mathbf{z}} \left(\mathbf{z}_{t,k} \mathbf{h}_{t-1,k}^{T} \right) + \sum_{t=1}^{T} P(k|\mathbf{x}_{t}) \boldsymbol{\Sigma}_{\mathbf{z},k}^{-1} \mathbf{A}_{k} E_{\mathbf{z}} \left(\mathbf{h}_{t-1,k} \mathbf{h}_{t-1,k}^{T} \right) = 0 \quad , \quad (14)$$

$$\Rightarrow \mathbf{A}_{k}^{new} = \sum_{t=1}^{T} P(k|\mathbf{x}_{t}) E_{\mathbf{z}} \left(\mathbf{z}_{t,k} \mathbf{h}_{t-1,k}^{T} \right) \left[\sum_{t=1}^{T} P(k|\mathbf{x}_{t}) E_{\mathbf{z}} \left(\mathbf{h}_{t-1,k} \mathbf{h}_{t-1,k}^{T} \right) \right]^{-1}$$

$$\frac{\partial \{ Q(\boldsymbol{\Theta} | \boldsymbol{\Theta}^{old}) \}}{\partial \mathbf{B}_{k}} = 0$$

$$\Rightarrow \frac{\partial \{ \sum_{t=1}^{T} P(k|\mathbf{x}_{t}) \left[-2E_{\mathbf{z}} (\mathbf{z}_{t,k}^{T}) \mathbf{B}_{k}^{T} \boldsymbol{\Sigma}_{\mathbf{x},k}^{-1} \mathbf{x}_{t,k} + E_{\mathbf{z}} \left(\mathbf{z}_{t,k}^{T} \mathbf{B}_{k}^{T} \boldsymbol{\Sigma}_{\mathbf{x},k}^{-1} \mathbf{B}_{k} \mathbf{z}_{t,k} \right) \right] \}}{\partial \mathbf{B}_{k}} = 0$$

$$\Rightarrow -\sum_{t=1}^{T} P(k|\mathbf{x}_{t}) \boldsymbol{\Sigma}_{\mathbf{x},k}^{-1} \mathbf{x}_{t,k} E_{\mathbf{z}} \left(\mathbf{z}_{t,k}^{T} \right) + \sum_{t=1}^{T} P(k|\mathbf{x}_{t}) \boldsymbol{\Sigma}_{\mathbf{x},k}^{-1} \mathbf{B}_{k} E_{\mathbf{z}} \left(\mathbf{z}_{t,k} \mathbf{z}_{t,k}^{T} \right) = 0 \quad (15)$$

$$\Rightarrow \mathbf{B}_{k}^{new} = \sum_{t=1}^{T} P(k|\mathbf{x}_{t}) \mathbf{x}_{t} E_{\mathbf{z}} \left(\mathbf{z}_{t,k}^{T} \right) \left[\sum_{t=1}^{T} P(k|\mathbf{x}_{t}) E_{\mathbf{z}} \left(\mathbf{z}_{t,k} \mathbf{z}_{t,k}^{T} \right) \right]^{-1}$$

$$\frac{\partial \{Q(\boldsymbol{\Theta}|\boldsymbol{\Theta}^{old})\}}{\partial \boldsymbol{\Sigma}_{\mathbf{z},k}} = 0$$

$$\Rightarrow \sum_{t=1}^{T} P(k|\mathbf{x}_{t}) \begin{bmatrix} \boldsymbol{\Sigma}_{\mathbf{z},k}^{-1} - \boldsymbol{\Sigma}_{\mathbf{z},k}^{-1} E_{\mathbf{z}}(\mathbf{z}_{t,k}\mathbf{z}_{t,k}^{T}) + \boldsymbol{\Sigma}_{\mathbf{z},k}^{-1} \mathbf{A}_{k} E_{\mathbf{z}}(\mathbf{h}_{t-1,k}\mathbf{z}_{t,k}^{T}) \mathbf{A}_{k}^{T} \boldsymbol{\Sigma}_{\mathbf{z},k}^{-1} \\ + \boldsymbol{\Sigma}_{\mathbf{z},k}^{-1} \mathbf{A}_{k} E_{\mathbf{z}}(\mathbf{z}_{t,k}\mathbf{z}_{t,k}^{T}) \mathbf{A}_{k}^{T} \boldsymbol{\Sigma}_{\mathbf{z},k}^{-1} \end{bmatrix} = 0 , \quad (16)$$

$$\Rightarrow \mathbf{\Sigma}_{\mathbf{z},k}^{new} = \frac{\sum_{t=1}^{T} P(k|\mathbf{x}_{t}) \left[E_{\mathbf{z}}(\mathbf{z}_{t,k}\mathbf{z}_{t,k}^{T}) - \mathbf{A}_{k}^{new} E_{\mathbf{z}}(\mathbf{h}_{t-1,k}\mathbf{z}_{t,k}^{T}) (\mathbf{A}_{k}^{new})^{T} + \mathbf{A}_{k}^{new} E_{\mathbf{z}}(\mathbf{z}_{t,k}\mathbf{z}_{t,k}^{T}) (\mathbf{A}_{k}^{new})^{T} \right]}{\sum_{t=1}^{T} P(k|\mathbf{x}_{t})}$$

$$\frac{\partial \{ Q(\boldsymbol{\Theta} | \boldsymbol{\Theta}^{old}) \}}{\partial \boldsymbol{\Sigma}_{\mathbf{z},k}} = 0$$

$$\Rightarrow \sum_{t=1}^{T} P(k | \mathbf{x}_{t}) \begin{bmatrix} \boldsymbol{\Sigma}_{\mathbf{x},k}^{-1} - \boldsymbol{\Sigma}_{\mathbf{x},k}^{-1} \mathbf{x}_{t} \mathbf{x}_{t}^{T} \boldsymbol{\Sigma}_{\mathbf{x},k}^{-1} + \boldsymbol{\Sigma}_{\mathbf{x},k}^{-1} \mathbf{B}_{k} E_{\mathbf{z}}(\mathbf{h}_{t,k}) \mathbf{x}_{t}^{T} \boldsymbol{\Sigma}_{\mathbf{x},k}^{-1} \\ + \boldsymbol{\Sigma}_{\mathbf{z},k}^{-1} \mathbf{B}_{k} E_{\mathbf{z}}(\mathbf{z}_{t,k} \mathbf{z}_{t,k}^{T}) \mathbf{B}_{k}^{T} \boldsymbol{\Sigma}_{\mathbf{z},k}^{-1} \end{bmatrix} = 0 \quad . \tag{17}$$

$$\Rightarrow \mathbf{\Sigma}_{\mathbf{x},k}^{new} = \frac{\sum_{t=1}^{T} P(k | \mathbf{x}_{t}) \left[\mathbf{x}_{t} \mathbf{x}_{t}^{T} - 2\mathbf{B}_{k}^{new} E_{\mathbf{z}}(\mathbf{h}_{t,k}) \mathbf{x}_{t}^{T} + \mathbf{B}_{k}^{new} E_{\mathbf{z}}(\mathbf{z}_{t,k} \mathbf{z}_{t,k}^{T}) (\mathbf{B}_{k}^{new})^{T} \right]}{\sum_{t=1}^{T} P(k | \mathbf{x}_{t})}$$

In order to update the factor coefficient P(k), since the factor involves more constraints, an optimization objective needs to be constructed separately for it. First, all terms related to P(k) are separated and represented as Equation (18).

$$g(k) = \sum_{t=1}^{T} \sum_{k=1}^{K} P(k|\mathbf{x}_t) InP(k).$$
 (18)

Due to the existence of constraints $\sum_{k=1}^{K} P(k) = 1$, the Lagrange multiplier is introduced and constructed into a Lagrange function form.

$$f(k) = g(k) + \lambda \left(\sum_{k=1}^{K} P(k) - 1\right),$$
 (19)

т

$$\sum_{t=1}^{T} P(k|\mathbf{x}_t) + \lambda P(k) = 0 \Rightarrow P(k) = -\frac{\sum_{t=1}^{L} P(k|\mathbf{x}_t)}{\lambda}.$$
(20)

We accumulate the *k* terms (for *k*) on both sides of Equations (19) and (20).

$$\left.\begin{array}{l} \sum\limits_{k=1}^{K} P(k) = -\frac{\sum\limits_{t=1}^{T} \sum\limits_{k=1}^{K} P(k|\mathbf{x}_{t})}{\lambda} \\ \sum\limits_{k=1}^{K} P(k|\mathbf{x}_{t}) = 1 \\ \sum\limits_{k=1}^{K} P(k) = 1 \end{array}\right\} \Rightarrow \lambda = -T.$$
(21)

We substitute the result back into Equation (20) to obtain the result of the factor coefficient.

$$P(k) = \frac{1}{T} \sum_{t=1}^{T} P(k|\mathbf{x}_t).$$
 (22)

We update the obtained parameter set $\Theta^{new} = \left\{ \mathbf{A}_{k}^{new}, \mathbf{B}_{k}^{new}, \mathbf{\mu}_{\pi,k}^{new}, \mathbf{\Sigma}_{\pi,k}^{new}, \mathbf{\Sigma}_{\mathbf{x},k}^{new}, P(k) | k = 1, 2, \cdots, K \right\}$, and iterate E and M steps to make the parameter matrix converge, that is, to satisfy Equation (23).

$$||\mathbf{\Theta}^{old} - \mathbf{\Theta}^{new}|| \le \sigma.$$
(23)

The optimal parameter set Θ is obtained. At this point, the parameter learning process of FDALM is over. The model is generated by training the data under normal operating conditions. Firstly, the normal data is pre-processed by normalization. Then the EM algorithm identifies the parameters. Finally, we obtain the trained FDALM. The flow chart of model training can be seen in Figure 4.



Figure 4. The flow chart of model training.

4. FDALM-Based Multimodal Process Monitoring Method

In order to avoid the segmentation of the process data leading to the destruction of the dynamic information in the sequence, it is necessary to make full use of the output of each modal factor in FDALM. However, the output of each factor mode is a representation of probability. If the construction statistics of each sub-mode are monitored separately, local false positives are prone to occur, and the overall monitoring effect cannot be achieved. Therefore, this chapter focuses on the use of Bayesian inference technology to fuse different factor modalities into the posterior failure probability of the final sample, so as to monitor the fluctuation of the whole process state in real time.

For a trained FDALM, the latent variable is the core force driving the operation of the system. Therefore, this variable contains important key state information of the modality. Although the latent variables of each sub-mode can establish the corresponding $T_{t,k}^2$ statistic.

$$T_{t,k}^{2} = E(\mathbf{z}_{t}^{q} | \mathbf{x}_{1:t}^{q})^{T} cov(\mathbf{Z}_{k})^{-1} E(\mathbf{z}_{t}^{q} | \mathbf{x}_{1:t}^{q}).$$
(24)

 Z_k is the set of latent variables in the *k*th mode in the training set, and $cov(Z_k)$ is the covariance of Z_k . However, different sub-models have their own control limits, and it is difficult to achieve the overall monitoring effect. In order to make full use of the sample posterior latent variable output in each mode, the monitoring results of each sub-mode are fused into the sample posterior failure probability by means of Bayesian inference method. First, we construct the probability of failure of the observation sample in the *k*th mode.

$$P_{T^{2}}^{k}\left(F\left|\mathbf{x}_{t}^{q}\right) = \frac{P_{T^{2}}^{k}\left(\mathbf{x}_{t}^{q}\right|F\right)P_{T^{2}}^{k}(F)}{P_{T^{2}}^{k}\left(\mathbf{x}_{t}^{q}\right)} \\ P_{T^{2}}^{k}\left(\mathbf{x}_{t}^{q}\right) = P_{T^{2}}^{k}\left(\mathbf{x}_{t}^{q}\left|N\right)P_{T^{2}}^{k}(N) + P_{T^{2}}^{k}\left(\mathbf{x}_{t}^{q}\left|F\right)P_{T^{2}}^{k}(F)\right)$$
(25)

Among them, $P_{T^2}^k(F)$ and $P_{T^2}^k(N)$ represent the prior probability of abnormal and normal process data, which can be skillfully combined with the significance level α .

$$P_{T^{2}}^{k}(F) = 1 - \alpha P_{T^{2}}^{k}(N) = \alpha$$
(26)

Among them, α is the significance level, and the actual size is the balance between false positives and false negatives. In order to obtain the failure probability $P_{T^2}^k\left(F | \mathbf{x}_t^q\right)$ of a new data sample, it is necessary to construct the conditional probability of the observed samples under normal and abnormal conditions. In this section, they are defined as Equation (27).

$$P_{T^{2}}^{k}(\mathbf{x}_{t}|N) = \exp\left\{-\frac{T_{t,k}^{2}}{T_{lim,k}^{2}}\right\}$$

$$P_{T^{2}}^{k}(\mathbf{x}_{t}|F) = \exp\left\{-\frac{T_{t,k}^{2}}{T_{t,k}^{2}}\right\}$$
(27)

Among them, $T^2_{\lim,k}$ is the control limit of each mode, and its value is uniquely determined by the degree of freedom *d* of the sub-model and the significance level α ; $T^2_{t,k}$ is the T^2 statistic of the *k*th sub-mode, and the calculation method is shown in Equation (24). After the probability of failure of each local model is determined in the observation data, Bayesian inference technology is used to further fuse the probability of failure of the sample in each sub-mode to obtain the posterior failure probability $T^2_{t,final}$ of the sample.

$$T_{t,final}^{2} = \sum_{k=1}^{K} P(k|\mathbf{x}_{t}) P_{T^{2}}^{k}(F|\mathbf{x}_{t}).$$
(28)

By comparing the posterior failure probability with the significance level α , it is judged

whether the system fails. The judgment logic is shown in Equation (29).

$$\begin{cases} T_{t,final}^2 \leq \alpha, & \text{normal} \\ T_{t,final}^2 > \alpha, & \text{abnormal} \end{cases}$$
(29)

The flow chart of process data modeling and monitoring is shown in Figure 4. Since the detection by constructing chi-square statistics for a single mode will result in a local false alarm, to improve the overall monitoring effect, this paper calculates the posterior failure probability of the sample in a certain mode. We then obtain the posterior failure probability of the sample by using Bayesian fusion methods. If the actual solitary probability of the sample exceeds the threshold, the monitoring is abnormal, and the control threshold is the average value between the failure rate and false alarm rate, which is generally taken as 0.05, i.e., tolerating 5% of false alarms.

Further, the main steps of the FDALM-based multimodal process monitoring method are shown in Algorithm 1:

Algorithm 1: Steps of an FDALM-based Process Monitoring Method

Input: In order to train the model parameters, the observation variables are selected from the process variables and quality variables, and the sampling sequence of *T* consecutive moments is selected as the training data set, denoted as $\mathbf{x}_{1:T}$; in order to test the monitoring performance, the sampling sequence of N consecutive moments is selected online as the Test data set, denoted as $\mathbf{x}_{1:N}$.

Step 1: The training and testing data sets $\{\mathbf{x}_{1:T}, \mathbf{x}_{1:N}\}$ are preprocessed by a normalization method. **Step 2:** Use the dynamic order identification algorithm and clustering algorithm to identify the order coefficient L and factor K of the model, determine the structure of FDALM, and initialize the model parameters Θ^{old} .

Step 3: Identification of FDALM Parameters Using Improved EM Algorithm.

Step 4: Determine the control threshold $T_{k,\text{lim}}^2$ and significance level α of the sub-model, calculate the latent variable distribution of each observation sample under each sub-model online, and use the Bayesian inference technology to fuse the output of each sub-mode into the sample posterior failure probability $T_{t,final}^2$.

Step 5: Compare $T_{t,final}^2$ with the significance level α to judge the state of the process. **Output:** Output system operating status.

5. Application Research on the Preparation Process of Ternary Cathode Material *5.1. Model Building*

To establish a reasonable mathematical model for the preparation process data, it is necessary to select the observed variables from the system process variables as the model input. Since the distribution of temperature field during sintering will have a significant impact on the performance of battery material, the specific manifestations are: over-burning will cause changes in material morphology and internal structure, and underburning will not provide sufficient activation energy for chemical reactions. By analyzing the mechanism of the sintering process, the decomposition reaction in the heating stage is an endothermic process that requires sufficient and suitable heat. Under-burning will lead to low water removal efficiency and affect subsequent oxidation reactions; overburning will lead to energy waste and increase production costs. Therefore, a reasonable temperature distribution is very important for the sintering process. By analyzing the actual production process data, the temperature sensors installed on the upper and lower sides of each temperature zone can indirectly reflect the temperature distribution of the sintering temperature zone. Reasonable temperature field distribution will help the production of high-quality products. At the same time, some process abnormal states are included in the product indicators. For example, the abnormality of the outlet replacement chamber will lead to an increase in the residual lithium content. In order to comprehensively monitor the common faults in the process, the temperature variables and residual lithium content of the entire temperature area are selected as the observation variables of the monitoring model.

In the actual production process, the main temperature anomalies are abnormal temperature increase/decrease and shutdown failures. Accidents such as the aging of the thermocouple and the fracture of the heating rod during the production process will cause the sintering temperature to rise abnormally. An abnormality such as the excessive current of the thermocouple and the large exhaust port can directly lead to an abnormal decrease of the temperature. The shutdown failure often occurs in the circuit short circuit, transmission jams, and other abnormalities. In order to verify the monitoring effect of the model under different faults, a total of 2200 consecutive time data were selected, and the sampling period was 30 min. The ternary cathode material roller kiln is generally stable during manufacturing and the temperature does not change much, so the sampling time in the plant is generally set as 30 min. All the data in this paper come from the industrial platform carried out in our laboratory. To simulate the continuity of the production process, all our faults are generated continuously, so the information at the time of continuous switching is retained to facilitate the ability of the model to model dynamically. It is not possible to put these different data segments together, because this would destroy the useful information between sequences, and it would be difficult to capture these dynamic changes when modeling dynamically, resulting in inaccurate modeling or a false alarm. The sequence data are collected under the same batch of battery material produced in the factory, and include the common process failures in the production process. In addition, the normalized mean and variance of the normalized data are used to normalize the training data and project their state distributions into a uniform space to facilitate the distinction between normal and abnormal states. The detailed physical description of this batch of data is shown in Table 1.

Table 1. Process data descriptions.

Туре	Data Description	Sampling Number	
Normal	This interval is normal data	1st-1000th	
Fault 1	1001st–1200th is normal data; 1201st–1400th is the abnormal temperature increase in the third temperature zone	1001st-1400th	
Fault 2	1401st–1600th is normal data; 1601st–1800th is the abnormal temperature drop in the third temperature zone	1401st-1800th	
Fault 3	1801st–2000th is normal data; 2001st–2200th is shutdown fault	1801st-2200th	

The product quality of the ternary cathode material is particularly sensitive to the sintering temperature. From a macro perspective, it is affected by the fluctuation of battery raw material and ambient temperature. The environment temperature in summer is $30 \,^\circ$ C and $10 \,^\circ$ C in winter. In the actual manufacturing, to overcome fluctuation in the environment temperature, it is necessary to adjust the operating parameters, thus changing the optimal sintering system, so that there are two optimal sintering systems in different environments temperature. The process data is collected for multimodal characteristics. The process data is clustered by an improved affine propagation clustering algorithm [29], which can adaptively determine the optimal parameters of the algorithm, thereby obtaining the optimal factor *K* of FDALM. The data in the normal data interval, that is, the 1st–1000th data, is used to classify the data set. The result shows that the data is divided into two categories. In order to visualize the division effect, the data is subjected to PCA dimension reduction, and the contribution is 99%. The first two-dimensional variables of the data are visualized, as shown in Figure 5.

It can be seen from the visualization result that the classification algorithm divides the process data into two categories. The results fully demonstrate that the data obeys multiple distributions, so the assumption of a single distribution does not hold. Accurate monitoring of this process requires further consideration of the multimodal nature of the process data.



Figure 5. Flow chart of the process monitoring method based on FDALM.

In order to establish a factor dynamic autoregressive latent variable model, the model structure parameters that need to be determined include order coefficient *L* and factor *K*. Among them, the order coefficient *L* is identified by the trend similarity algorithm L = 3; the physical interpretation of the factor *K* is the number of types of data division, which is obtained by the improved affine propagation clustering algorithm, K = 2. In order to verify the rationality of factor selection, this section designs experiments to establish FDALM under different factor *K* and compare its monitoring effect. The normal sample interval 1st~1200th is selected as the training data set, and the fault 1 data set is used to test the monitoring effect of the model. In order to evaluate the actual monitoring effect of the model. In each compare the compare the fault Detection Rate (*FDR*), are used.

$$FAR = \frac{N_{FAR}}{N_n}$$

$$FDR = \frac{N_{FDR}}{N_n}.$$
(30)

Among them, N_{FAR} represents the number of normal samples misclassified by the monitored method, and N_n is the number of all normal samples. N_{FDR} is the number of fault samples correctly detected by the monitoring method, and N_f is the number of all fault samples. Table 2 shows the monitoring effect of the method proposed in this section under different factor *K*.

Table 2. FDALM monitoring results for different K.

K	1	2	3
FAR	$0.150 \\ 0.945$	0.020	0.055
FDR		0.965	0.940

Table 2 shows the monitoring results indicators of FDALM under the factor K of (1, 2, 3). It can be seen that when factor K = 2, the method can obtain better monitoring results by combining the monitoring results and training costs. At the same time, Figure 6 shows the monitoring diagram of the monitoring method under different factors. It can be seen that choosing an appropriate K is helpful to smooth the monitoring results, and the monitoring results also verify the effectiveness of the classification algorithm. Finally, the factor K is selected as 2.



Figure 6. Process data classification diagram.

5.2. Process Monitoring

In order to compare the monitoring effect of the proposed monitoring method (FDALM) on dynamic multimode processes, two other monitoring methods are established for comparison in this section, which are representatives of static models: Probabilistic Partial Least Squares Regression (PPLSR) [30], and the representative of dynamic models: dynamic autoregressive latent variable model (DALM). Table 3 presents the *FAR* and *FDR* indicators of the monitoring results of the three monitoring methods in the data sets with different fault types.

Type –	PPI	LSR	DA	LM	FDA	ALM
	FAR	FDR	FAR	FDR	FAR	FDR
Fault 1	0.240	0.120	0.120	0.925	0.020	0.965
Fault 2	0.100	0.350	0.090	0.970	0.045	0.985
Fault 3	0.315	0.980	0.085	1.000	0.040	1.000

Table 3. Monitoring results for different faults.

From the final monitoring effect, FDALM is better than the DALM method, and DALM is better than the PPLSR method. Considering the structure of the model, the PPLSR model, as a static modeling method, only performs effective dimensionality reduction on the data set without considering the changes in the data sequence. Although DALM considers the dynamic information between sequences, it is difficult to adapt to monitoring of multimode processes; the FDALM model considers the dynamics and multimode characteristics of process data as a whole, so it is more comprehensive. In order to more intuitively show the monitoring performance at different times, Figures 7–10 further present the monitoring results of the three models under different faults.

Compared with the traditional PPLSR, FDALM introduces both dynamic and multimodal information of the process data, which can better describe the dynamic changes and modal fluctuations of the sintering process. Therefore, the characterization of the process is more comprehensive. In addition, the advantages of FDLAM are also reflected in the modeling and fusion technology. First, the establishment of FDALM for process data can more accurately describe the dynamic and multimode characteristics of the process; it is prone to local false positives, and, finally, modeling sequence data does not lose inter-sequence information.



Figure 7. FDALM results at different K.



Figure 8. Monitoring results of fault 1.



Figure 9. Monitoring results of fault 2.

Compared with the single DALM monitoring method, the introduction of factor technology makes the modeling method suitable for multimode processes, which further improves the monitoring effect. The intuitive performance is that the parameters of the model in different modes of the process are different. Through factor weighting and Bayesian inference technology, the models established in different sub-modes can not only capture the modal information, but also establish local information for the fusion of the overall failure probability, so it can better follow the system dynamics and modal



fluctuations. Compared with DALM, the overall fault false alarm rate and detection rate of the multimodal feature monitoring method are improved by 48.2% and 1.9%, respectively.

Figure 10. Monitoring results of fault 3.

To sum up, the multimodal process monitoring method based on FDALM simultaneously solves the problems of dynamics, randomness, and multimodality of process data in the battery sintering process. It has a strong ability to describe complex production processes, and the monitoring results can play a better role in predicting the actual production process.

6. Conclusions

In this paper, a multimode process monitoring method based on FDALM was proposed. Firstly, in the technology of dynamic autoregressive latent variable model, with the help of the factor modeling method, DALM suitable for single mode was extended to multimode form, so that the model can consider the dynamic and multimode characteristics of process data at the same time. We introduced the use of the EM algorithm to learn the model parameters. Next, the failure probability of the sample under each sub-model was fused into the posterior failure probability of the sample by means of Bayesian inference technology to realize the process output of each factor model. Finally, the simulation results compared with other models show that the proposed monitoring method was able to track the modal fluctuations of the process.

An important issue in industrial process monitoring applications is the problem of multiple sampling rates. The method proposed in this paper assumes that the input and output data have the same sampling rate. If the sampling rate is inconsistent, some data is removed by down sampling. However, a more worthwhile approach is to combine semi-supervised learning methods, which can improve data utilization by training data on imbalanced input and output data. Another practical problem is the nonlinear relationships between process data, which are common in industrial processes. How to effectively deal with this issue deserves further research in the near future to make monitoring methods more applicable.

Author Contributions: Conceptualization, N.C., W.G. and C.Y.; methodology, F.H. and J.C.; software, F.H. and K.W.; validation, J.C.; writing—review and editing, N.C., F.H., J.C. and K.W.; super-vision, N.C. and K.W.; project administration, W.G. and C.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded in part by the Key Program of National Natural Science Foundation of China (62033014) and in part by the Application Projects of Integrated Standardization and New Paradigm for Intelligent Manufacturing from the Ministry of Industry and Information Technology of China in 2016 and in part by the Fundamental Research Funds for the Central Universities of Central South University (2021zzts0700).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available upon request from the first author. The data are not publicly available due to intellectual property protection.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Venkatasubramanian, V.; Rengaswamy, R.; Kavuri, S.N. A review of process fault detection and diagnosis Part II: Quantitative model and search strategies. *Comput. Chem. Eng.* 2003, 27, 313–326. [CrossRef]
- Venkatasubramanian, V.; Rengaswamy, R.; Kavuri, S.N.; Yin, K. A review of process fault detection and diagnosis: Part III: Process history based methods. *Comput. Chem. Eng.* 2003, 27, 327–346. [CrossRef]
- 3. Venkatasubramanian, V.; Rengaswamy, R.; Yin, K.; Kavuri, S.N. A review of process fault detection and diagnosis: Part I: Quantitative model-based methods. *Comput. Chem. Eng.* **2003**, *27*, 293–311. [CrossRef]
- 4. Qin, S.J. Statistical process monitoring: Basics and beyond. J. Chemom. 2003, 17, 480–502. [CrossRef]
- 5. Kano, M.; Hasebe, S.; Hashimoto, I.; Ohno, H. A new multivariate statistical process monitoring method using principal component analysis. *Comput. Chem. Eng.* 2001, 25, 1103–1113. [CrossRef]
- Wang, X.; Kruger, U.; Lennox, B. Recursive partial least squares algorithms for monitoring complex industrial processes. *Control Eng. Pract.* 2003, 11, 613–632. [CrossRef]
- Chen, H.; Jiang, B.; Ding, S.X.; Huang, B. Data-Driven Fault Diagnosis for Traction Systems in High-Speed Trains: A Survey, Challenges, and Perspectives. *IEEE Trans. Intell. Transp. Syst.* 2020, 23, 1700–1716. [CrossRef]
- Kruger, U.; Zhou, Y.; Irwin, G.W. Improved principal component monitoring of large-scale processes. J. Process Control 2004, 14, 879–888. [CrossRef]
- 9. Kruger, U.; Kumar, S.; Littler, T. Improved principal component monitoring using the local approach. *Automatica* 2007, 43, 1532–1542. [CrossRef]
- 10. Foka, Y.; Prosper, H.B.; Brambilla, N.; Kovalenko, V. Deep Learning and Bayesian Methods. EPJ Web Conf. 2017, 137, 11007.
- 11. Ge, Z.; Chen, X. Dynamic Probabilistic Latent Variable Model for Process Data Modeling and Regression Application. *IEEE Trans. Control. Syst. Technol.* **2019**, *27*, 323–331. [CrossRef]
- 12. Barber, D. Bayesian Reasoning and Machine Learning; Cambridge University Press: Cambridge, UK, 2012.
- Chen, N.; Hu, F.H.; Chen, J.Y.; Chen, Z.W.; Gui, W.H.; Li, X. A Process Monitoring Method Based on Dynamic Autoregressive Latent Variable Model and Its Application in the Sintering Process of Ternary Cathode Materials. *Machines* 2021, 9, 229. [CrossRef]
- Ng, Y.S.; Srinivasan, R. An adjoined multi-model approach for monitoring batch and transient operations. *Comput. Chem. Eng.* 2009, 33, 887–902. [CrossRef]
- 15. Yoo, C.K.; Villez, K.; Lee, I.-B.; Rosén, C.; Vanrolleghem, P. Multi-model statistical process monitoring and diagnosis of a sequencing batch reactor. *Biotechnol. Bioeng.* 2007, *96*, 687–701. [CrossRef]
- 16. Ben Khediri, I.; Weihs, C.; Limam, M. Kernel k-means clustering based local support vector domain description fault detection of multimodal processes. *Expert Syst. Appl.* **2012**, *39*, 2166–2171. [CrossRef]
- 17. Zhao, Z.G.; Liu, F. A new method for process monitoring based on mixture probabilistic principal component analysis models. In Proceedings of the Third International Symposium on Neural Networks, Chengdu, China, 28–31 May 2006; pp. 939–944.
- 18. Zhao, S.J.; Zhang, J.; Xu, Y.M. Monitoring of processes with multiple operating modes through multiple principle component analysis models. *Ind. Eng. Chem. Res.* **2004**, *43*, 7025–7035. [CrossRef]
- 19. Zhao, S.J.; Zhang, J.; Xu, Y.M. Performance monitoring of processes with multiple operating modes through multiple PLS models. *J. Process Control* **2006**, *16*, 763–772. [CrossRef]
- 20. Yu, J. A new fault diagnosis method of multimode processes using Bayesian inference based Gaussian mixture contribution decomposition. *Eng. Appl. Artif. Intell.* **2013**, *26*, 456–466. [CrossRef]
- 21. Yu, J.; Qin, S.J. Multimode process monitoring with Bayesian inference-based finite Gaussian mixture models. *AIChE J.* **2008**, *54*, 1811–1829. [CrossRef]
- 22. Ge, Z.; Song, Z. Mixture Bayesian regularization method of PPCA for multimode process monitoring. *AIChE J.* **2010**, *56*, 2838–2849. [CrossRef]
- Wang, F.; Tan, S.; Yang, Y.; Shi, H. Hidden Markov Model-Based Fault Detection Approach for a Multimode Process. *Ind. Eng. Chem. Res.* 2016, 55, 4613–4621. [CrossRef]
- Wang, J.; Shao, W.; Song, Z. Student's-t Mixture Regression-Based Robust Soft Sensor Development for Multimode Industrial Processes. Sensors 2018, 18, 3968. [CrossRef] [PubMed]
- 25. Huang, X.-K.; Tian, X.-Y.; Zhong, Q.; He, S.-W.; Huo, C.-B.; Cao, Y.; Tong, Z.-Q.; Li, D.-C. Real-time process control of powder bed fusion by monitoring dynamic temperature field. *Adv. Manuf.* **2020**, *8*, 380–391. [CrossRef]
- 26. Chen, J.; Gui, W.; Dai, J.; Jiang, Z.; Chen, N.; Li, X. A hybrid model combining mechanism with semi-supervised learning and its application for temperature prediction in roller hearth kiln. *J. Process. Control.* **2021**, *98*, 18–29. [CrossRef]

- 27. Egorova, E.; Rudakova, I.; Rusinov, L.; Vorobjev, N. Diagnostics of sintering processes on the basis of PCA and two-level neural network model. *J. Chemom.* 2018, 32, e2959. [CrossRef]
- 28. Panić, B.; Klemenc, J.; Nagode, M. Improved Initialization of the EM Algorithm for Mixture Model Parameter Estimation. *Mathematics* **2020**, *8*, 373. [CrossRef]
- 29. Fu, X.; Yu, Z. Multiple models soft-sensing method based on improved adapt affinity propagation. *Comput. Appl. Chem.* **2016**, *33*, 111–116.
- Wang, Z.Y.; Liang, J. A JITL-Based Probabilistic Principal Component Analysis for Online Monitoring of Nonlinear Processes. J. Chem. Eng. Jpn. 2018, 51, 874–889. [CrossRef]